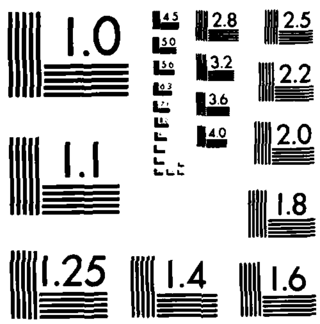
AD-A144 992 ACHIEVING GENERALITY OVER CONDITIONS: COMBINING THE MULTITRAIT MULTIMETHO..(U) COLORADO UNIV AT BOULDER CENTER FOR RESEARCH ON JUDGMENT AND P.. 1/1

UNCLASSIFIED K R HAMMOND ET AL. 02 AUG 84 CRJP-255 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ACHIEVING GENERALITY OVER CONDITIONS:
COMBINING THE MULTITRAIT MULTIMETHOD MATRIX
AND THE REPRESENTATIVE DESIGN OF EXPERIMENTS

Kenneth R. Hammond, Robert M. Hamm,
and Janet Grassia

ACHIEVING GENERALITY OVER CONDITIONS: COMBINING THE MULTITRAIT MULTIMETHOD MATRIX AND THE REPRESENTATIVE DESIGN OF EXPERIMENTS

Kenneth R. Hammond, Robert M. Hamm,
and Janet Grassia

Center for Research on Judgment and Policy

University of Colorado

Boulder, Colorado 80309

Report No. 255

August 1984

This research was supported by the Engineering Psychology Programs, Office of Naval Research, Contract N00014-81-C-0591, Work Unit Number NR 197-073 and by BRSG Grant #RR07013-14 awarded by the Biomedical Research Support Program, Division of Research Resources, NIH. Center for Research on Judgment and Policy, Institute of Cognitive Science, University of Colorado.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CRJP 255	2. GOVT ACCESSION NO. AD-A144 992	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Achieving Generality over Conditions: Combining the Multitrait Multimethod Matrix and the Representative Design of Experiments		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kenneth R. Hammond and Robert M. Hamm		8. CONTRACT OR GRANT NUMBER(s) N00014-81-C-0591
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Research on Judgment and Policy Institute of Cognitive Science University of Colorado, Boulder, CO 80309		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-073
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE August 2, 1984
		13. NUMBER OF PAGES 73
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: Distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

multitrait-multimethod, representative design, expert judgment, generalization, methodology

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Campbell and Fiske's (1959) multitrait multimethod matrix and Brunswik's (1956) representative design of experiments are combined and extended in order to increase our ability to generalize over conditions in both experimental psychology and the study of individual differences. A study of expert judgment illustrates the application of Campbell and Fiske's methodology to an experimental design involving multiple concepts and methods, and in addition, criterion measures for

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

block 20

each concept. The criterion measures make it possible to complement Campbell and Fiske's internal validity matrix with an external validity matrix. Inclusion of the correlations among criteria in the external validity matrix is consistent with Brunswik's argument that generalization over conditions depends upon the representation of ecological relations among experimental conditions. Procedures are described for computing measures of convergent and discriminant validity for each matrix and for combining the data from both matrices.

ACHIEVING GENERALITY OVER CONDITIONS: COMBINING THE MULTITRAIT MULTIMETHOD MATRIX AND THE REPRESENTATIVE DESIGN OF EXPERIMENTS

Doubts about the generality of results produced by psychological research have been expressed with increasing frequency since Koch observed, after a monumental review of scientific psychology in 1959, that there is "a stubborn refusal of psychological findings to yield to empirical generalization" (1959, pp. 729-788). Brunswik (1952, 1956), Campbell and Stanley (1966), Cronbach (1975), Epstein (1979, 1980), Einhorn and Hogarth (1981), Greenwald (1975, 1976), Hammond (1966), Meehl (1978) and Simon (1979) among others, have also called attention to this situation and some (Epstein, 1980; Greenwald, 1976) have referred to it as a "crisis." All regard it as a fundamental, persistent problem in psychological research.

In an effort to develop a methodology that will provide generality without the loss of rigor, we build upon two previous methodological suggestions, (a) the multitrait multimethod matrix introduced by Campbell and Fiske (1959) and (b) the representative design of experiments introduced by Brunswik (1956). Data from a study of experts who were required to employ three modes of cognition in each of three judgment tasks (see Appendix A) provided a unique opportunity not only to make use of the multitrait multimethod matrix, but to extend it.

In their 1959 study of the field of individual differences, Campbell and Fiske convincingly demonstrated the faults of the conventional single-concept single-operation methodology. The overwhelming majority of studies they examined showed that results were more likely to be determined by the methods employed by the experimenters than by the traits hypothesized to account for the results. Although they showed that this failure to separate the effects of operation (method) from the effects of concept (trait) can be both <u>demonstrated</u> and <u>avoided</u> by use of the multitrait multimethod matrix, there has been little change in conventional research methodology.

The problem is not that Campbell and Fiske's work went unrecognized. It became a milestone in the methodological literature of psychology, and by 1983 had been cited over 1000 times. Yet in spite of the potential of the multitrait multimethod matrix for breaking the grip of a simpleminded operationism on psychological research, the method is for the most part simply not used. Presumably researchers have avoided it for tactical reasons, since it introduces conceptual complexity (which concepts and which methods should be compared?) and requires considerable additional labor and apparatus within a single study. Or perhaps there is general unawareness of the ephemeral character of results produced by single-concept single-method operationism. Whatever the reason, among tens of thousands of studies of individual differences, Turner (cited in Fiske, 1981) found only 70 published matrices between 1967 and 1980 (see Fiske, 1981, for a general review).

The multitrait multimethod matrix has probably never been used in experimental psychology, although its logic is equally applicable to that field (cf. Fiske, 1981). We examined the 62 articles in Volume 9 (1983) of the Journal of Experimental Psychology: Human Perception and Performance to ascertain whether researchers currently make a systematic effort to separate method variance from concept variance. The persistence of one-concept one-method operationism was evident: only 18 articles were found to employ more than one concept or more than one method; and of these, only four used more than one concept and more than one method. None, however, systematically separated method variance from concept variance; only one of the authors indicated cognizance of this methodological requirement. The multitrait multimethod approach was never mentioned.

In parallel fashion, Brunswik's (1943, 1952, 1956) argument that generalization over conditions requires the representation of ecological conditions in the design of experiments must be considered a milestone in the methodological literature of psychology; his work, too, has been cited over 1000 times, yet representative designs are seldom employed (see Hammond & Wascoe, 1980, for some examples). Representative design was never mentioned in the 62 articles examined in the volume cited above. The same reasons that led students of individual differences to forgo the use of the multitrait multimethod matrix also lead experimental psychologists to forgo the use of representative design; both are more difficult and time-consuming to execute than standard laboratory experiments.

Plan of the Article

In what follows we first present a description of the Campbell/Fiske internal validity matrix; second, indicate our extension of it to an external validity matrix that incorporat: the theory of representative design of experiments; third, show the complementarity of the two matrices; and fourth, illustrate how both matrices can be used to achieve generalization over conditions.

The Campbell-Fiske Internal Validity Matrix

The internal validity multitrait multimethod matrix, presented in Table 1, is developed from a set of test scores taken from a group of subjects (Campbell & Fiske, 1959). The scores for each subject are correlated over several traits and methods. The authors describe the matrix as follows:

> This illustration involves three different traits, each measured by three methods, generating nine separate variables. It will be convenient to have labels for various regions of the matrix, and such have been provided in Table [1]. The reliabilities will be spoken of in term of three reliability diagonals, one for each method. The reliabilities could also be designated as the monotrait-monomethod values. Adjacent to each reliability diagonal is the heterotrait-monomethod triangle. The reliability diagonal and the adjacent heterotrait-monomethod triangle make up a monomethod block. A heteromethod block is made up of a validity diagonal (which could also be designated as monotrait-heteromethod values) and the two

heterotrait-heteromethod triangles lying on each side of it. Note that these two heterotrait-heteromethod triangles are not identical.

In terms of this diagram, four aspects bear upon the question of validity. In the first place, the entries in the validity diagonal should be significantly different from zero and sufficiently large to encourage further examination of validity. This requirement is evidence of convergent validity. Second, a validity diagonal value should be higher than the values lying in its column and row in the heterotrait-heteromethod triangles. That is, a validity value for a variable should be higher than the correlations obtained between that variable and any other variables having neither trait nor method in common. This requirement may seem so minimal and so obvious as to not need stating, yet an inspection of the literature shows that it is frequently not met, and may not be met even when the validity coefficients are of substantial size. In Table [1], all the validity values meet this requirement. A third common-sense desideratum is that a variable correlate higher with an independent effort to measure the same trait than the measures designed to get at different traits which happen to employ the same method. For a given variable, this involves comparing its values in the validity diagonals with its values in the heterotrait-monomethod triangles. For variables A1, B1, and C1, this requirement is met to some degree. A fourth desideratum is that the same pattern of trait interrelationship be shown in all of the heterotrait triangles of both the monomethod and

> heteromethod blocks. The hypothetical data in Table [1] meet this requirement to a very marked degree, in spite of the different general levels of correlation involved in the several heterotrait triangles. The last three criteria provide evidence for discriminant validity. (1959, pp. 82-83).

The value of this methodology is indisputable, and its application will yield definite and useful conclusions regarding the validity of psychological traits or theoretical concepts in general (see, e.g., Brewer & Collins, 1981; Fiske, 1981). The results from such a matrix will have populational and task generality insofar as the trait domain, the apparatus/method domain and the subject domain have been adequately sampled. The results, therefore, speak to the question of the construct validity of the traits investigated separate from the methods used, within the restraints chosen by the investigator.

Insert Table 1 about here

Extension of the Campbell/Fiske Approach

Campbell and Fiske (1959) developed the multitrait multimethod matrix in order to evaluate the (a) internal validity of certain (b) traits within the study of (c) individual differences based on (d) group data. We extend their method by (a) adding an external validity matrix; (b) using both the internal and the external validity matrices to evaluate concepts in general instead of traits; (c) using both matrices to test propositions, in the tradition of experimental psychology; (d) making the behavior of the individual rather than of the group the fundamental unit of analysis, although group data can be analyzed as well. (See Hammond, McClelland, &

Mumpower, 1980, pp. 115-127 on the advantages of single-subject analysis; also Meehl, 1978, on the deficiencies of conventional between-group and within-group analyses.)

The External Validity Matrix

Table 2 presents an external validity matrix that is based upon correlations between nine sets of engineers' judgments, made under three methods (cognitive modes) for each of three concepts, and three criteria. The three validity diagonals contain monoconcept correlations between each set of judgments (one for each method) and the criterion of the same concept against which the judgments are compared. The triangles consist of heteroconcept correlations between the judgments made in each condition (concept-method unit) and the criterion for a different concept. A method block consists of a validity diagonal and the heteroconcept triangles on either side of it.

The coefficients in the external validity matrix in Table 2 are different from those in the internal validity matrix in that each correlation in the external validity matrix is between judgments and measures of a criterion rather than between two responses. Aside from this very important difference, the interpretation of the coefficients with respect to the questions of convergent and discriminant validity is quite similar. As in the internal validity matrix, correlations in the external validity diagonal that are sufficiently large are evidence of convergent validity. In Table 2 the coefficients in the diagonals within each method block would show the external convergent validity of the judgment of each concept by that method. Comparison of the average of these diagonal values across the three concepts would indicate the relative external convergent

validity of each method. The heteroconcept triangles consist of the correlations of the expert's judgments of one concept (by a particular method) with the criterion measure of a different concept. Evidence of discriminant validity exists when a value in a validity diagonal is higher than the values lying in its column and row in the heteroconcept triangles. Further tests of external discriminant validity are described below.

Insert Table 2 about here

The External Validity Matrix and the Representative Design of Experiments

The argument for the representative design of experiments is explicated in the external validity matrix because the naturally occurring intercorrelations among criterion variables are represented in the matrix (see Table 2). For example, if the correlation between criteria for c1 and c2 in Table 2 were .5, we would expect all correlations between judgments of c1 and the criterion for c2 (and vice versa) to be as high as but no higher than .5 *if* an engineer is performing appropriately. The intercorrelations among the criteria, or intraecological correlations, thus provide a *standard* for the heteroconcept correlations in the external validity matrix, and in the internal validity matrix as well. Without ecological representativeness as a standard, all such intercorrelations are changed by the experimenter to zero in the conventional systematic design of experiments. Therefore, generalization cannot be achieved on logical grounds, and indeed is not achieved empirically, as the psychologists cited above emphasized.

Complementarity of the Internal and External Validity Matrices: Evaluating Coherence, Performance, and Competence

The usefulness of analyzing the external validity matrix in conjunction with Campbell and Fiske's internal validity matrix is that the information provided by these matrices is complementary and makes possible an evaluation of cognitive coherence, performance, and competence. The distinction between coherence and performance is intended to parallel the traditional distinction between the coherence and correspondence theories of truth (see, e.g., White, 1967 and Prior, 1967). The coherence theory focuses on the extent to which statements of facts or judgments put forward cohere (or "hang together") with one another, that is, are related by logical implication. The internal validity matrix parallels the coherence theory of truth in the sense that it demands logical rather than external, empirical justification. Although the internal matrix does include empirical, factual material, no reference to empirical criteria outside the matrix itself is required to establish the internal validity of a set of psychological concepts. All that is required is that a logical criterion be met, namely, that convergent validities should be high and discriminant validities should be low.

The correspondence theory of truth, on the other hand, is concerned with the extent to which our beliefs about the world perform, or correspond, with respect to independently determined facts. Therefore an independent measure of the concepts in question is required in order to test the correspondence between what a theory predicts and what exists. The external validity matrix thus parallels the correspondence theory of truth in that it demands the evaluation of the empirical correspondence between psychological concepts and some independent measure of them.

Finally, because both matrices can be developed for a single subject (as we demonstrate below), it is possible to combine the results from each matrix into a single measure to provide a higher order indicator of each expert's judgment that we shall call "competence" (see also McClelland, 1973). Since we derive the measure of competence from measures of coherence and performance that are based on variations in both method and concept, our derivation copes directly with the problem of generalization. In the present case, for example, the conclusions about an expert's coherence and performance, and thus competence, are clearly based on, and thus limited to, his/her behavior over the three methods and three concepts employed in the study.

Summary of Similarities and Differences between Campbell and Fiske (1959) and the Present Approach

The two efforts are similar in that each provides comparisons of convergent validities and discriminant validities across concepts and methods (see Tables 1 and 2); but there are several differences. First, the internal validity matrix does not include test-criterion relations, but the external validity matrix does. Therefore it contains correlation coefficients that indicate the relation between measures of each subject's behavior and external, empirical criteria. As a result, the meaning of the entries in the cells is different in the two matrices. The correlation coefficient in each cell in the Campbell/Fiske internal validity matrix indicates the correlation between pairs of test measures, whereas the correlation coefficients in the external validity matrix indicate the correlation between a behavioral measure and an external criterion.

Second, the role of the individual subject in the two kinds of analysis is very different. Each correlation coefficient in the Campbell and Fiske (1959) multitrait multimethod matrix is across individuals, while in a multiconcept multimethod analysis each is across the objects of judgment, within a single individual. More specifically, in a multitrait multimethod analysis, each of n individuals is measured on j (traits) times k (methods) occasions, and one multitrait multimethod matrix is made for the whole set of individuals. In a multiconcept multimethod analysis, each of n individuals judges each of p objects on j (concepts) times k (methods) occasions, and a separate multiconcept multimethod matrix is constructed for each of the n individuals.

Third, because the external validity matrix must contain at least two criterion variables in order to separate concept from method, the relations between criteria in circumstances toward which the generalization is intended must be measured and taken into consideration when the subject's performance is evaluated. Conventional experimental psychology has been able to sidestep this matter only because of its persistent, implicit acceptance of single-concept single-method operationism. It is precisely at this point, however, that the external validity matrix is directly linked to Brunswik's (1956) representative design of experiments. In representative designs, intra-ecological correlations between criteria cannot be ignored and arbitrarily set to zero as is customary. This convention introduces a design feature that must, and has, frustrated generalization of results because the results are obtained under conditions seldom if ever present in the conditions of application. The use of representative design, however, means that correlations among criterion variables in the experiment will represent those in the circumstances to

which the results of the experiment are intended to generalize, or apply. In short, the same logic of inductive inference that we apply when generalizing from subject sample to subject population will apply to generalizing from experimental conditions to any other set of conditions (see, for example, Brunswik, 1943, 1952, 1956; Hammond, 1966; Hammond & Wascoe, 1980; Einhorn & Hogarth, 1981; Epstein, 1979, 1980).

Fourth is a difference in aims. Campbell and Fiske's principal aim was to enhance our methodological ability to evaluate the construct validity of traits (Cronbach & Meehl, 1955). We take that aim to have been achieved in principle (no one has challenged it), if not in practice. We aim therefore to build upon that achievement by showing that both matrices can be applied to experimental psychology as well as to the study of individual differences (cf. Cronbach, 1975). In addition, we intend to show that the internal validity matrix and the external validity matrix provide complementary information: (a) the internal validity matrix method can be used to evaluate the coherence of an expert's judgments, (b) the external validity matrix can be used to evaluate the performance of an expert's judgments, and (c) measures of coherence and performance can be combined to provide a measure of competence.

Illustrative Application

The Use of the Internal Validity Matrix in a Study of Expert Judgment

Data for an internal validity matrix based on a study of 20 highway engineers' judgments of the concepts of aesthetics, safety and capacity using intuitive, quasi-rational, and analytical methods (see Appendix A) are presented in Table 3. The data for the matrix were generated from the

mean of the 20 engineers' judgments for each of the 40 highways presented to them for each concept-method pair. Thus, the matrix illustrates the particulars of the behavior of an artificial engineer constructed from the mean judgments of this group. Data from the artificial engineer are presented mainly to illustrate the use of the method; no inferences can be drawn from the matrix in Table 3 to a matrix generated by any one engineer. Illustrations of individual matrices are provided below.

Insert Table 3 about here

Each of the descriptions of the matrix presented by Campbell and Fiske (1959) apply to the matrix in Table 3. The three validity diagonals contain values that are high, relative to the heteroconcept triangles adjacent to them, thus providing evidence for internal convergent and discriminant validity.

Use of the External Validity Matrix in a Study of Expert Judgment

Table 4 presents the artificial engineer's external validity matrix, also based on the mean of 20 engineers' judgments.

Insert Table 4 about here

Convergent validity of concepts. The external validity coefficient for the artificial engineer's aesthetics judgments made by the film strip method is .855, by the bar graph method is .945, and by the formula method is .951, thus producing a mean external convergent validity value across all three methods of .926 for aesthetic judgments. (Note: Fisher's z-transformation is used in the calculation of mean values.) Averaging

validity correlations pertaining to safety from the three method boxes, the mean convergent validity is .568; similarly, averaging the judgment-criterion correlations for capacity produces a mean convergent validity value of .530. In short, the data suggest that, irrespective of the method used, the artificial engineer judged highway aesthetics more accurately than highway safety or capacity, and judged safety and capacity with equal accuracy.

Convergent validity of methods. A measure of the external convergent validity for each method may be calculated by averaging the judgment-criterion correlations within each of the diagonals (.86, .70, .29; .95, .68, .83; .95, .23, .27), thus obtaining external validities for each method (.67, .85, .65). These results suggest that the artificial engineer judged these three concepts most accurately in the quasi-rational mode. Finally, the mean of the latter three coefficients is .74. This measure is informative because it may be used to compare one group of experts with another, to compare one individual with another (in the case when a matrix is constructed for each individual), or to evaluate the effect of a change in condition in either case. Moreover, the referential domain of this measure is clear; it is general over the three methods and three concepts employed in the study, as well as the group of engineers selected.

Measuring discriminant validity with reference to intra-ecological correlations. The intra-ecological correlations among empirical measures of the concepts permit an additional method for assessing discriminant validity. The correlation between the criterion measures of the concepts provides a standard against which to compare the heteroconcept correlations

between the expert's judgment of one concept and the criterion measure of a different concept. For example, if the correlation between aesthetics and safety is -.275, then it is appropriate for an engineer's judgments of aesthetics to be correlated -.275 with safety (see Appendix B). Similarly, if the correlation between two criterion measures is low (as for safety and capacity, .180), then the heteroconcept correlations should also be low. In short, the observed correlations between judgments of aesthetics, safety and capacity for an engineer are not to be compared to a standard of zero (an arbitrary demand for complete independence regardless of task conditions) but to a standard that is representative of task conditions, if we are properly to evaluate the discriminant validity of the judgments of these concepts with these methods.

To "untie" these variables, in other words to force zero intercorrelations among them, is (a) to invite the engineer to judge an unrepresentative set of conditions and thus (b) to extrapolate his results illegitimately from irrelevant conditions to the relevant ones. These two tactics have an embarrassingly long history in psychology; they are customarily explained away by arguments that "this is the best we can do" and/or "it doesn't matter, anyway." Neither argument is correct, but neither is necessary; the external validity form of the multiconcept multimethod matrix makes it possible to evaluate the competence of experts (or other subjects) in relation to the task conditions to which their judgments are to be applied.

The examples presented below illustrate the detailed application of both the internal and external validity matrices to the study of expert judgment. The first section describes the use of both matrices for testing

propositions in the context of experimental psychology, and the second describes the use of the matrices in connection with the study of individual differences.

Application to Experimental Psychology

Internal Validity Matrix

The analyses to be reported in this section require that a matrix, similar to that for the artificial engineer of Table 3, above, be produced for each engineer, and that convergent or discriminant validities be determined for each.

It is possible to derive one criterion of convergent validity and four criteria of discriminant validity from the internal validity matrix. The criterion for convergent validity and one for discriminant validity are described below. The remaining criteria for internal discriminant validity are described in Appendix C.

Convergent validity. The convergent validity measure (monoconcept heteromethod correlations between judgments of the same concept using different methods) can be used to test hypotheses concerning the empirical status of each concept. For example,

H1: Each theoretical concept has empirical meaning, i.e., there is convergent validity for each concept across methods and within an appropriate sample of subjects.

Hypothesis 1 can be tested by asking whether, for each subject, judgments of the quantity of a concept covary, independently of the methods used to make the judgments. For example, for the artificial engineer (Table 3) the correlation between the film strip and bar graph methods for the aesthetics concept is .890; for the film strip and formula methods, .864; and for the bar graph and formula methods, .985. The overall convergent validity for aesthetics is the mean of these correlations (z-transformed), .938, which is significant at $p < .001$. A matrix was developed for each of the 20 engineers individually, and this procedure was carried out for each of the three concepts. All 20 engineers had significant positive convergent validities for aesthetics, 16 for safety, and 17 for capacity. Hence we conclude that each of the three concepts is capable of being measured by appropriate subjects independently of the method used; generality has been achieved over three methods.

More specific hypotheses may also be addressed. for example,

H2: No concept has higher or lower convergent validity than any other.

To test Hypothesis 2, the computed mean of the z-transforms of the three aesthetics convergent validities (indicated in the previous paragraph) is compared to the means of the safety and capacity convergent validities, for each engineer. The results indicate that 17 of the 20 engineers had greatest convergent validity when judging the aesthetics concept (Chi-squared = 21.75, $p < .001$). A t-test analysis shows that, over the 20 engineers, the convergent validity score for aesthetics was significantly higher than the score for safety ($t = 5.08$, $p < .001$) and for

capacity ($t = 5.66$, $p < .001$). Again, the generality of the results is not contingent upon a single method; the domain of generality over concepts, methods and subjects is made evident.

Questions regarding the relative efficacy of methods over concepts may also be addressed. For example,

> H3: No method pair has higher or lower convergent validity than any other method pair.

To test Hypothesis 3, we must consider the convergent validities related to each pair of methods. When the film strip and bar graph methods are applied to aesthetics, the convergent validity is .890, to safety, .713, and to capacity, .591 for the artificial engineer of Table 3. The mean (via z-transforms) of these correlations is .761. The mean for each of the possible method pairs is calculated through the development of a matrix for each engineer, and the order among pairs is determined, similar to the analysis used for testing Hypothesis 2. For 17 of the 20 engineers the bar graph and formula were the method pair that produced the highest convergent validity across the three concepts (Chi-squared = 21.754, $p < .001$). This result tells us which pair of methods across the three concepts is best for achieving convergent validity with regard to these three concepts.

Discriminant validity. Convergent validity informs us about the covariance of judgments across methods, and thus about the status of a concept independent of the method used to measure it. In addition, however, we need to know whether the concept is discriminable from other

proposed theoretical entities. The first internal discriminant validity analysis employed in the examples below compares monoconcept heteromethod correlations to heteroconcept heteromethod correlations. Campbell and Fiske (1959) gave first priority to this test; for although many people would think it "so minimal and obvious as not to need stating," (p. 82) they observed that it often fails to be true. We therefore illustrate the test for the following hypothesis:

H4: All pairs of concepts are equally discriminable.

This hypothesis will be tested by calculating an index for each concept pair for each engineer, and looking for evidence of any concept being more, or less, discriminable than the others, for a statistically significant number of engineers. To illustrate the calculation of the index for the aesthetic and safety concepts, for the artificial engineer of Table 3, we compare the correlations from the validity (monoconcept heteromethod) diagonals that involve either aesthetics (.890, .864, .985) or safety (.713, .393, .422) with the correlations from the heteroconcept heteromethod triangles that involve both concepts (.283, .244, .360, .093, .548, and .209). (The sign on all heteroconcept correlations involving aesthetics was reversed because the intra-ecological correlations between the criterion measures of aesthetics and safety, and of aesthetics and capacity, were negative.) In order to aggregate these comparisons into an index, we subtract the mean of the z-transformations of the second set of correlations (.306) from the mean of the z-transformations of the first set (1.156), which produces an index (.850) of the discriminability of the aesthetics and safety concepts. The corresponding index for aesthetics and

capacity is .913; for safety and capacity, -.047. Thus, for the artificial engineer aesthetics and capacity are the easiest concepts to discriminate, and safety and capacity are most difficult to discriminate (a result which carries some practical implications).

This index of discriminant validity is calculated for each concept pair from each subject's matrix, and the order among concept pairs is determined. For all 20 engineers, the safety and capacity concepts were least discriminable (Chi-squared = 37.053, $p < .001$). Therefore null hypothesis 4 is rejected, for the engineers' judgments of safety and capacity are more similar to each other than either is to their judgment of aesthetics. The remaining three indices of internal discriminant validity are described in Appendix C.

External Validity Analysis

One measure of convergent validity and three measures of discriminant validity can be derived from the external validity matrix. In addition, two measures of external discriminant validity can be produced using data from the internal validity matrix.

Convergent validity. The external convergent validity measure is based on the correlation between the engineer's judgments of a concept and the criterion measure of that concept. We examine first the relative convergent validity of each concept, thus:

H5: No concept has higher or lower external convergent validity than any other.

Hypothesis 5 is tested by averaging the z-transforms of the correlations for each concept across methods, and then comparing the averages for each concept. Thus the aesthetics concept had higher convergent validity than safety or capacity for all 20 engineers (Chi-squared = 37.053, $p < .001$). Despite the counterintuitive nature of this result, it has a claim to our attention; it is general across three methods and stands against two other concepts.

Similar questions of external convergent validity can be addressed to methods. For example,

> H6: No method has higher or lower external convergent validity than any other.

Hypothesis 6 is tested by averaging the z-transforms of the correlations for each method across concepts, and comparing methods. The film strip method was found to have the lowest convergent validity for 18 of the 20 engineers (Chi-squared = 26.404, $p < .001$). It is least dependable in the context of this study. Methods and results for Hypotheses 5 and 6 are given in more detail in Hammond, Hamm, Grassia, and Pearson (1984).

Discriminant validity. The external validity matrix provides three ways of measuring external discriminant validity, and two additional measures can be produced from the internal validity matrix in combination with the criterion intercorrelations. These measures can be used to ask whether concepts can be discriminated accurately. For example,

H7: All pairs of concepts are equally discriminable.

The first external discriminant validity measure is analogous to the first internal discriminant validity measure, and is used to test Hypothesis 7 just as the latter was used to test Hypothesis 4: by calculating an index of discriminability for each concept pair for each engineer, comparing the concept pairs, and determining whether any particular order among the concept pairs occurred in a significant number of engineers. Thus, for the artificial engineer in Table 4, the discriminability of the aesthetics and safety concepts is measured by subtracting the mean of the z-transforms of the heteroconcept correlations involving aesthetics and safety (-.016, .362, .233, .497, .313, and .226) from the mean of the z-transformations of the achievement correlations involving aesthetics <u>or</u> safety (aesthetics: .855, .945, .951; safety: .702, .683, .226), a difference of .855. This figure is calculated for each concept pair for each engineer; the safety and capacity concepts were least discriminable for each of the 20 engineers (Chi-squared = 37.05, $p < .001$), a result that is consistent with that obtained in the internal validity matrices.

The availability of information about the intercorrelation among the measured criteria makes possible four additional procedures besides the first measure of external discriminant validity described above. The second and third procedures involve direct comparison of heteroconcept correlations with the correlations between the criterion measures of the two concepts, for the external and internal validity matrices respectively. The fourth and fifth procedures involve testing, for both matrices, whether the pattern of correlations in each heteroconcept triangle is identical to the pattern of correlations among the three criterion measures.

The second and third external discriminant validity measures allow us to ask whether there is systematic over- or underdiscrimination between concepts by testing the following hypothesis:

H8: Concepts are discriminated accurately.

The second external discriminant validity measure, which compares heteroconcept correlations from the external validity matrix with the corresponding correlations between the criterion measures, was used to test Hypothesis 8. A parallel test could be carried out with the third external discriminant validity index, which uses heteroconcept correlations from the internal validity matrix.

From each heteroconcept correlation in the external validity matrix, the corresponding criterion intercorrelation (intra-ecological correlation) is subtracted (after z-transformation). The mean of these differences for the set of heteroconcept correlations corresponding to a pair of concepts indicates the extent of the engineer's under- or overdiscrimination of the concepts. This procedure can be carried out for the safety and capacity concepts for the artificial engineer (Table 4). We subtract the z-transform of .180, the correlation between their criterion measures, from the mean of the z-transforms of the heteroconcept correlations involving aesthetics and safety (.683, .399, .516, .437, .199, .383), producing a difference of .302. The positive sign of this number indicates that, overall, the artificial engineer underdiscriminates between safety and capacity (confirming two prior results). This procedure was carried out for each concept pair, for each engineer. Fourteen of the 20 overdiscriminated between aesthetics and safety (Chi-squared = 2.45,

df = 1, NS), 15 over discriminated between aesthetics and capacity (Chi-squared = 4.05, $p < .05$), and 19 underdiscriminated between safety and capacity (Chi-squared = 14.45, $p < .001$).

The testing of a hypothesis using the fifth operation measuring external discriminant validity is described in Appendix C.

Summary

In this section we have illustrated the application of the multiconcept multimethod validity analysis to topics typically of concern to experimental psychologists: testing theoretical propositions regarding the comparison of concepts and methods. This was done by using the internal validity matrix, which is concerned solely with the relations among different judgments of the concepts, obtained under different methods; and with the external validity matrix, concerned with the relation between the judgments and the criterion measures of the concepts.

Our illustration highlights the complementarity of these two analyses. We found in both the internal and external validity analyses that the aesthetics concept has the highest convergent validity; that the best pair of methods to use to obtain discriminant validity (in these conditions) is the quasi-rational, bar graph method and the analytical, formula-producing method; and that safety and capacity are least discriminable from each other. The external validity analysis was able to put this last finding in sharper perspective than could the internal validity analysis. It showed that the engineers underdiscriminate safety and capacity in comparison with the intercorrelation between the criterion measures of these concepts. The engineers' judgments of these two concepts are most highly correlated,

while in fact the criterion measures of these concepts have the lowest intercorrelation. This could have been otherwise; that is, if safety and capacity had actually been very highly correlated, the engineers might have overdiscriminated them even if the internal validity analysis had indicated that these two concepts are discriminated less than any other pairs of concepts. The external validity analysis provides the only way to determine which of these possibilities is true.

Application to Individual Differences

The multiconcept multimethod approach can be used to study individual differences in the competence of expert judgment. The need for individual comparisons is apparent from Tables 5, 6, 7, and 8, which show the internal and external validity matrices for two engineers. Engineer A's validity correlations are high (mean of 9 monoconcept heteromethod correlations from internal validity matrix = .705; mean of 9 monoconcept correlations from external validity matrix = .741), while Engineer B's are low (mean from internal matrix = .358; mean from external matrix = .609). Similar differences can be seen in their discriminant validities.

Insert Tables 5, 6, 7, 8 about here

Indices of Coherence, Performance and Competence

Individual differences among engineers can be studied using numerical measures of convergent and discriminant validity derived from both the internal and external validity matrices. A procedure for evaluating validity can be converted into a numerical measure by adding (or subtracting, as appropriate) the means of the z-transforms of the

correlations in all the relevant cells in the matrix. The formulas for producing these indices are given in Table 9 and explained in Appendix D.

These measures can be combined into indices that measure overall internal validity (see Figure 1), which indicates the coherence of the engineer's judgments; the corresponding index of external validity indicates the engineer's performance, i.e., the correspondence between his judgments and reality. And the mean of these two indices provides a measure of the engineer's overall competence. Each index can be produced at different levels of aggregation (e.g., for each concept or for each pair of methods; see columns of Table 9), thus allowing numerical comparisons among these indices at each level.

Insert Figure 1 and Table 9 about here

Measurements of coherence and performance are of special theoretical importance. The coherence of a person's judgments is the central characteristic of one of the traditional theories of knowledge, the coherence theory of truth. And the performance of a person's judgments is the central characteristic of a second traditional theory of knowledge, the correspondence theory of truth. Therefore, taken together, indices of internal and external validity inform us about a person's competence in the context of two historic theories of truth.

The methodology described here makes it possible to measure coherence and performance over several concepts and methods. Thus the generality of the behavior of each subject is explicated in terms of each theory of knowledge in the context of a different matrix of concepts and methods, each of which provides its own methodological justification for the generalization of results.

Among the experts in the example used here, a fairly high relation (.60) was found between coherence and performance. The treatment of coherence and performance as cognitive traits thus will allow us to examine empirically theoretical questions of importance to both philosophers and psychologists. For example:

1. Should competence always be a joint product of coherence and performance? Should these traits always be additive? Or is coherence a necessary but not a sufficient condition for performance? Common sense suggests that this should be so. But the relation between these traits may depend upon the complexity of the material and the degree of intellectual training required to master it. That is, variation in competence in, say, atomic physics may produce a very high correlation between coherence and performance, whereas variation in competence in financial forecasting may not. In short, coherence and performance may combine in different ways to provide competence, depending upon the nature of the material to be dealt with and the degree of training of the subject.

2. How should the measures of coherence and performance be combined into an overall measure of competence? Should they be weighted according to their relative importance and/or the quality of the measures? Common practice is to consider these measures separately. Moreover, different approaches to the study of cognition give greater consideration to one or the other of these aspects of competence. Studies within the framework of artificial intelligence and problem solving, for example, weight the experts'

coherence (and the coherence of the computer program that simulates the expert) very highly while placing less weight on performance. Judgment and decision researchers do the opposite (see Hammond, 1983). Explicating the concept of competence in terms of coherence and performance thus suggests that these two currently independent fields of research are investigating complementary aspects of competence among experts.

Comparison of the Competence of the Individual Experts and the Artificial Expert

Tables 3 and 4 (above) show the data for the artificially constructed engineer, produced by taking the mean of all the engineers' judgments of each highway, within each of the nine cells, and then performing a multiconcept multimethod analysis on these data. Would such an artificial expert, built upon aggregated judgments, provide more competent judgments than the individual experts?

Table 10 contrasts the validity indices and subindices for the artificial engineer with the corresponding indices for the lowest, mean, and best of the individual engineers. For all indices the artificial engineer's validity indices were better than the mean of the individual engineers' indices. Most important, for two indices (internal and external convergent validity) the artificial engineer's index was better than that of the best engineer. Finally, combining engineers' individual judgments produced judgments that were more competent than all but one engineer's judgments.

Insert Table 10 about here

Summary

Individual differences were found in the quality of experts' judgment. Numerical measures were created for a number of procedures for measuring internal and external convergent and discriminant validity. These were combined into indices for the internal validity matrix (pertaining to the coherence of experts' judgment) and for the external validity matrix (pertaining to their performance). A correlation of .60 between coherence and performance was found among the engineers used in the illustrative example. The coherence and performance of the artificial engineer, created by averaging all individual engineers' judgments of each condition of the study, proved superior to that of the individual engineers.

Discussion

As several noted psychologists have observed, psychological research lacks the cumulative character critical to the development of a science. In any such circumstance suspicion would arise that the scientific discipline in question is the captive of a flawed theoretical or methodological dogma. Since theories are numerous in psychology, but methodology is uniform throughout graduate schools and journal reviews, dogmatic methodology must be the prime suspect.

In an attempt to address the methodological problem of generalization we have extended and integrated the pioneering efforts of Campbell and Fiske (1959) and Brunswik (1956). Using individual experts' judgments of the safety, capacity and aesthetics of highways made under three

conditions, we first created a multiconcept multimethod matrix of internal validity for the judgment of concepts about highways, using different methods of eliciting judgments. This contrasts with Campbell and Fiske's multitrait multimethod matrix for the measurement of traits of persons, using different trait-measuring methods. Second, we used criterion measures for the concepts to create an external validity matrix. Measures of convergent and discriminant validity can be calculated from each of these matrices and used to address questions concerning, for example, how easily concepts can be discriminated or how well each method works. Taking full cognizance of the empirical relations among criteria in the determination of external discriminant validity conforms to Brunswik's demand for the representative design of experiments. Because the intercorrelations among the concepts are taken into account, the domain of the generality of the results is explicit.

The logic of the multiconcept multimethod matrix is based on what Feigl called "triangulation in logical space" (Feigl, 1958; see also Campbell & Fiske, 1959, p. 84). From a logical point of view, the methods and concepts selected for study should be completely independent; the "triangulation" should approximate a right triangle as nearly as possible. Thus, Campbell and Fiske (1959) discuss "convergence of the independent methods" and cite Cronbach and Meehl's argument that the use of "diverse criteria give[s] greater weight to the claim of construct validity than do. . .predictions of very similar behavior" (Cronbach & Meehl, 1955, p. 295). Brunswik, however, emphasized the fact that the ecological variables that so often serve as criteria for psychologists' concepts are not independent, i.e., orthogonal to one another. Therefore, from the researcher's point of view, Fiegl's concept of "triangulation in logical

space" is not to be seen as a goal, but as a condition that serves didactic purposes, without regard to the demands of specific problems. The proper goal for the researcher (in contrast to the logician) is "triangulation in empirical space," in which the logician's worship of orthogonality is replaced by the researcher's worship of generalization. Informative as the logician's remarks undoubtedly are, the proper goal of basic research is generalization of results; and that goal can best be achieved through the use of "representative triangulation," in experiments as well as in studies of individual differences.

Addendum

Curiously, the literature of modern physics does not seem to include many treatises on methodological issues relating to reliability and validity of experiments, although there is a long history of treatises on measurement in physics (also apparent in psychology). A recent paper (Franklin & Howson, 1984) entitled "Why do scientists prefer to vary their experiments?" treats this topic as a contemporary one, thus suggesting that it does not have a long history (the oldest topical reference is 1979). Also, there appears to be no systematic treatment in physics of the problem of separation of method from concept such as carried out by Campbell and Fiske (1959). Personal communication with Allan Franklin confirms this conclusion. If psychology and physics are indeed beginning to recognize a common methodological problem of considerable importance, much might be gained from a joint consideration of "why scientists prefer to vary their experiments" (although it is not at all clear that <u>all</u> scientists do).

A comparison of the manner in which various experimental (physics, chemistry, biology) and nonexperimental (astronomy, archeology) disciplines treat the matter of repetition of experiments, the separation of reliability and validity, and/or the separation of concept from method is beyond the scope of the present article. Nevertheless, it is worth mentioning that our impression is that Campbell and Fiske's (1959) contribution, based on Feigl's (1958) original work, provides a more sophisticated, detailed examination of this matter than exists elsewhere (cf. Hacking, 1983).

References

Brewer, M., & Collins, B. (1981). Scientific inquiry and the social sciences. San Francisco: Jossey-Bass.

Brunswik, E. (1943). Organismic achievement and environmental probability. Psychological Review, 50, 255-272.

Brunswik, E. (1952). The conceptual framework of psychology. In International encyclopedia of unified science (Vol. 1, No. 10). Chicago: University of Chicago Press.

Brunswik, E. (1956). Perception and the representative design of psychological experiments (2nd ed.). Berkeley: University of California Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validiation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Campbell, D. T., & Stanley, J. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand-McNally

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Delucchi, K. L. (1983). The use and misuse of Chi-square: Lewis and Burke revisited. Psychological Bulletin, 94, 166-176.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 32, 53-88.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. Journal of Personality and Social Psychology, 37, 1097-1126.

Epstein, S. (1980). The stability of behavior II. Implications for psychological research. American Psychologist, 35, 790-806.

Feigl, H. (1958). The mental and the physical. In H. Feigl, M. Scriven, and G. Maxwell (Eds.), Minnesota studies in the philosophy of science (Vol. 2): Concepts, theories and the mind-body problem. Minneapolis: University of Minnesota Press, pp. 370-497.

Fiske, D. W. (1981). Problems with language imprecision. San Francisco: Jossey-Bass.

Franklin, A., & Howson, C. (1984). Why do scientists prefer to vary their experiments? Studies in History and Philosophy of Science, 15, 51-62.

Greenwald, A. G. (1975). Significance, nonsignificance, and interpretation of an ESP experiment. Journal of Experimental Social Psychology, 11, 180-191.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? _Psychological Bulletin_, _83_, 314-320.

Hacking, I. (1983). _Representing and intervening: Introductory topics in the philosophy of natural science_. New York: Cambridge University Press.

Hammond, K. R. (Ed.). (1966). _The psychology of Egon Brunswik_. New York: Holt, Rinehart, & Winston.

Hammond, K. R. (1983). _Teaching the new biology: Potential contributions from research in cognition_. In C. P. Friedman & E. F. Purcell (Eds.), _The new biology and medical education_. New York: Josiah Macy, Jr. Foundation, pp. 53-64.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1984). _The relative efficacy of intuitive and analytical cognition: A second direct comparison_ (Tech. Rep. No. 252). Boulder: University of Colorado, Center for Research on Judgment and Policy.

Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). _Human judgment and decision making: Theories, methods, and procedures_. New York: Praeger.

Hammond, K. R., & Wascoe, N. E. (Eds.). (1980). _Realizations of Brunswik's representative design_. San Francisco: Jossey-Bass.

Hays, W. L. (1973). _Statistics for the social sciences_ (2nd ed.). New York: Holt, Rinehart and Winston.

Highway Research Board. (1965). Highway Capacity Manual 1965 (Highway Research Board Special Report No. 87, National Research Council Publication 1328). Washington, DC: National Research Council.

Koch, S. (1959). Epilogue. In S. Koch (Ed.), Psychology: A study of a science (Vol. 3, pp. 729-788). New York: McGraw-Hill.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence." American Psychologist, 28, 1-14.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Prior, A. N. (1967). Correspondence theory of truth. In P. Edwards (Ed.), The encyclopedia of philosophy: Vol. 2 (pp. 223-232). New York: Macmillan.

Simon, H. A. (1979). Models of thought. New Haven: Yale University Press.

White, A. R. (1967). Coherence theory of truth. In P. Edwards (Ed.), The encyclopedia of philosophy: Vol. 2 (pp. 130-132). New York: Macmillan.

Footnotes

In constructing the internal validity matrix, repeated judgment reliabilities were not available from the data. Therefore they were estimated, using R from the linear best fit model of the engineer's judgments for the film strip and bar graph methods, and using the correlation between the judgments produced by corrected and uncorrected formulas for the formula method. Further details of these measures are available in Hammond, Hamm, Grassia, and Pearson (1984).

To determine whether the z-transformation of a correlation is significantly different from a zeta of zero (the expected correlation under the null hypothesis), the z-transformation is converted to a z-score by the formula (z-score minus zeta) divided by the variance of zeta (square root of [1/(N - 3)]), and the probability of the z-score is determined from tables for the normal distribution. See Hays (1973, p. 662).

Table 1

A Synthetic Multitrait-Multimethod Matrix

		Method 1			Method 2			Method 3		
	Traits	A_1	B_1	C_1	A_2	B_2	C_2	A_3	B_3	C_3
Method 1	A_1	(.89)								
	B_1	.51	(.89)							
	C_1	.38	.37	(.76)						
Method 2	A_2	.57	.22	.09	(.93)					
	B_2	.22	.57	.10	.68	(.94)				
	C_2	.11	.11	.46	.59	.58	(.84)			
Method 3	A_3	.56	.22	.11	.67	.42	.33	(.94)		
	B_3	.23	.57	.12	.43	.66	.34	.67	(.92)	
	C_3	.11	.11	.45	.34	.32	.58	.58	.60	(.85)

Table 2

External Validity Matrix

		Method 1			Method 2			Method 3		
	Concepts	C1	C2	C3	C1	C2	C3	C1	C2	C3
CRITERIA	C1	.855	-.362	-.473	.945	-.479	-.311	.951	-.226	.002
	C2	.016	.702	.683	-.233	.683	.516	-.313	.226	.199
	C3	-.172	.399	.291	-.298	.437	.833	-.316	.383	.266

Key:

monoconcept (validity)

heteroconcept

Table 3

Internal Validity Multiconcept Multimethod Matrix for Artificial Engineer

METHOD										
		Film Strip I			Bar Graph Q			Formula A		
	Concepts	E	S	C	E	S	C	E	S	C
I	E	(.957)								
	S	.035	(.825)							
	C	-.117	.951	(.793)						
Q	E	.890	-.283	-.364	(.996)					
	S	-.244	.713	.712	-.491	(.968)				
	C	-.051	.706	.591	-.282	.681	(.979)			
A	E	.864	-.360	-.452	.985	-.548	-.337	(.999)		
	S	-.093	.393	.427	-.209	.422	.452	-.232	(.415)	
	C	.176	.403	.333	.015	.270	.340	.004	.001	(.3[illegible])

Key:

( ) monoconcept monomethod (reliability)

monoconcept heteromethod (validity)

heteroconcept monomethod

heteroconcept heteromethod

Key:

E : Aesthetics

S : Safety

C : Capacity

Table 4

<u>External Validity Multiconcept Multimethod Matrix for Artificial Engineer</u>

	METHOD								
	Film Strip I			Bar Graph Q			Formula A		
Concepts	E	S	C	E	S	C	E	S	C
CRITERIA E	.855	-.362	-.473	.945	-.479	-.311	.951	-.226	.002
S	.016	.702	.683	-.233	.683	.516	-.313	.226	.199
C	-.172	.399	.291	-.298	.437	.833	-.316	.383	.266

Key:

monoconcept (validity)

heteroconcept

Key:

E : Aesthetics

S : Safety

C : Capacity

Table 5

Internal Validity Matrix for Engineer A

METHOD		Concepts	Film Strip I: E	Film Strip I: S	Film Strip I: C	Bar Graph Q: E	Bar Graph Q: S	Bar Graph Q: C	Formula A: E	Formula A: S	Formula A: C
METHOD	I	E	(.893)								
	I	S	.126	(.756)							
	I	C	.001	.536	(.549)						
	Q	E	.793	-.103	-.128	(.955)					
	Q	S	-.145	.467	.449	-.190	(.895)				
	Q	C	-.010	.144	.335	-.103	.470	(.893)			
	A	E	.786	-.223	-.153	.909	-.359	-.095	(1.000)		
	A	S	-.217	.557	.490	-.281	.875	.440	-.473	(1.000)	
	A	C	-.172	.435	.389	-.335	.604	.756	-.375	.616	(.994)

Key:

( ) monoconcept monomethod (reliability)

monoconcept heteromethod (validity)

heteroconcept monomethod

heteroconcept heteromethod

Key:

E : Aesthetics

S : Safety

C : Capacity

Table 6

External Validity Matrix for Engineer A

		METHOD								
		Film Strip I			Bar Graph Q			Formula A		
	Concepts	E	S	C	E	S	C	E	S	C
CRITERIA	E	.747	-.279	-.309	.843	-.286	-.101	.925	-.435	-.317
	S	.047	.583	.345	-.037	.538	.184	-.222	.691	.378
	C	-.194	.300	.268	-.326	.474	.769	-.358	.498	.873

Key:

monoconcept (validity)

heteroconcept

Key:

E : Aesthetics

S : Safety

C : Capacity

Table 7

Internal Validity Matrix for Engineer B

			METHOD								
			Film Strip I			Bar Graph Q			Formula A		
		Concepts	E	S	C	E	S	C	E	S	C
METHOD	I	E	(.716)								
		S	.355	(.694)							
		C	.511	.788	(.723)						
	Q	E	-.005	-.409	-.381	(.882)					
		S	-.094	.287	.332	-.267	(.717)				
		C	.092	.572	.448	-.461	.533	(.899)			
	A	E	.148	-.307	-.343	.770	-.366	-.375	(.997)		
		S	.126	.431	.467	-.541	.505	.585	-.483	(.921)	
		C	.138	.155	.058	-.258	-.087	.344	-.028	.120	(.663)

Key:

( ) monoconcept monomethod (reliability)

monoconcept heteromethod (validity)

heteroconcept monomethod

heteroconcept heteromethod

Key:

E : Aesthetics

S : Safety

C : Capacity

Table 8

External Validity Matrix for Engineer B

		METHOD								
		Film Strip I			Bar Graph Q			Formula A		
	Concepts	E	S	C	E	S	C	E	S	C
CRITERIA	E	.253	-.377	-.364	.742	-.449	-.461	.878	-.469	-.098
	S	.331	.500	.529	-.261	.499	.336	-.393	.518	-.092
	C	.195	.475	.260	-.364	.216	.783	-.161	.487	.668

Key:

monoconcept (validity)

heteroconcept

Key:

E : Aesthetics

S : Safety

C : Capacity

Table 9

<u>Measures of Validity</u>

		For Each Concept-Method Unit (m,j)	For Each Concept (m)	For Each Concept Pair (m,n) m≠n
INTERNAL VALIDITY	Reliability	$r_{m_j m_j}$	$\mathop{M}_{j} r_{m_j m_j}$	
	Convergent Validity	$\mathop{M}_{\substack{k\\k\neq j}} r_{m_j m_k}$	$\mathop{M}_{\substack{j,k\\j\neq k}} r_{m_j m_k}$	
	Discriminant Validity IDV1	$\mathop{M}_{\substack{k\\k\neq j}} r_{m_j m_k} - \mathop{M}_{\substack{k\\k\neq j}} \mathop{M}_{\substack{n\\n\neq m}} r_{m_j n_k}$	$\mathop{M}_{\substack{j,k\\j\neq k}} r_{m_j m_k} - \mathop{M}_{\substack{j,k\\j\neq k}} \mathop{M}_{\substack{n\\n\neq m}} \dfrac{\left(r_{m_j n_k} + r_{m_k n_j}\right)}{2}$	$\mathop{M}_{\substack{j,k\\j\neq k}} \dfrac{\left(r_{m_j m_k} + r_{n_j n_k}\right)}{2} - \mathop{M}_{\substack{j,k\\j\neq k}} \dfrac{\left(r_{m_j n_k} + r_{m_k n_j}\right)}{2}$
	Discriminant Validity IDV2	$r_{m_j m_j} - \mathop{M}_{\substack{n\\n\neq m}} r_{m_j n_j}$	$\mathop{M}_{j} r_{m_j m_j} - \mathop{M}_{j} \mathop{M}_{\substack{n\\n\neq m}} r_{m_j n_j}$	$\mathop{M}_{j} \dfrac{\left(r_{m_j m_j} + r_{n_j n_j}\right)}{2} - \mathop{M}_{j} r_{m_j n_j}$
	Discriminant Validity IDV3	$\mathop{M}_{\substack{k\\k\neq j}} r_{m_j m_k} - \mathop{M}_{\substack{n\\n\neq m}} r_{m_j n_j}$	$\mathop{M}_{\substack{j,k\\j\neq k}} r_{m_j m_k} - \mathop{M}_{j} \mathop{M}_{\substack{n\\n\neq m}} r_{m_j n_j}$	$\mathop{M}_{\substack{j,k\\j\neq k}} \dfrac{\left(r_{m_j m_k} + r_{n_j n_k}\right)}{2} - \mathop{M}_{j} r_{m_j n_j}$
EXTERNAL VALIDITY	Convergent Validity	$r_{m\,m_j}$	$\mathop{M}_{j} r_{m\,m_j}$	
	Discriminant Validity EDV1	$^{*}r_{m\,m_j} - \mathop{M}_{\substack{n\\n\neq m}} \dfrac{\left(r_{m\,n_j} + r_{n\,m_j}\right)}{2}$	$\mathop{M}_{j} r_{m\,m_j} - \mathop{M}_{j} \mathop{M}_{\substack{n\\n\neq m}} \dfrac{\left(r_{m\,n_j} + r_{n\,m_j}\right)}{2}$	$\mathop{M}_{j} \dfrac{\left(r_{m\,m_j} + r_{n\,n_j}\right)}{2} - \mathop{M}_{j} \dfrac{\left(r_{m\,n_j} + r_{n\,m_j}\right)}{2}$
	Discriminant Validity EDV2	$\mathop{M}_{\substack{n\\n\neq m}} \left\vert r_{m\,n_j} - r_{m\,n}\right\vert$	$\mathop{M}_{j} \mathop{M}_{\substack{n\\n\neq m}} \dfrac{\left(\left\vert r_{m\,n_j} - r_{m\,n}\right\vert + \left\vert r_{n\,m_j} - r_{m\,n}\right\vert\right)}{2}$	$\mathop{M}_{j} \dfrac{\left(\left\vert r_{m\,n_j} - r_{m\,n}\right\vert + \left\vert r_{n\,m_j} - r_{m\,n}\right\vert\right)}{2}$
	Discriminant Validity EDV3	$\mathop{M}_{k} \mathop{M}_{\substack{n\\n\neq m}} \left\vert r_{m_j n_k} - r_{mn}\right\vert$	$\mathop{M}_{j,k} \mathop{M}_{\substack{n\\m\neq n}} \dfrac{\left(\left\vert r_{m_j n_k} - r_{m\,n}\right\vert + \left\vert r_{m_k n_j} - r_{m\,n}\right\vert\right)}{2}$	$\mathop{M}_{j,k} \dfrac{\left(\left\vert r_{m_j n_k} - r_{m\,n}\right\vert + \left\vert r_{m_k n_j} - r_{m\,n}\right\vert\right)}{2}$

*Alternatively: $r_{m\,m_j} - \frac{1}{2} \mathop{M}_{\substack{n\\n\neq m}} \left(r_{m\,n_j} + r_{n\,m_j}\right)$

and similarly for all terms that are divided by two.

Table 9 (continued)

			For Each Method (j)	For Each Method Pair (j,k) j≠k	Overall
INTERNAL VALIDITY	Reliability		$\underset{m}{M}\; r_{m_j m_j}$		$\underset{j}{M}\;\underset{m}{M}\; r_{m_j m_j}$
	Convergent Validity		$\underset{\substack{k\\k\neq j}}{M}\;\underset{m}{M}\; r_{m_j m_k}$	$\underset{m}{M}\; r_{m_j m_k}$	$\underset{\substack{j,k\\j\neq k}}{M}\;\underset{m}{M}\; r_{m_j m_k}$
	Discriminant Validity	IDV1	$\underset{\substack{k\\k\neq j}}{M}\;\underset{m}{M}\; r_{m_j m_k} - \underset{\substack{k\\k\neq j}}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{r_{m_j n_k} + r_{m_k n_j}}{2}\right)$	$\underset{m}{M}\; r_{m_j m_k} - \underset{\substack{m,n\\m\neq n}}{M}\left(\frac{r_{m_j n_k} + r_{m_k n_j}}{2}\right)$	$\underset{\substack{j,k\\j\neq k}}{M}\;\underset{m}{M}\; r_{m_j m_k} - \underset{\substack{j,k\\j\neq k}}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{r_{m_j n_k} + r_{m_k n_j}}{2}\right)$
		IDV2	$\underset{m}{M}\; r_{m_j m_j} - \underset{\substack{m,n\\m\neq n}}{M}\; r_{m_j n_j}$		$\underset{j}{M}\;\underset{m}{M}\; r_{m_j m_j} - \underset{j}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\; r_{m_j n_j}$
		IDV3	$\underset{\substack{k\\k\neq j}}{M}\;\underset{m}{M}\; r_{m_j m_k} - \underset{\substack{m,n\\m\neq n}}{M}\; r_{m_j n_j}$	$\underset{m}{M}\; r_{m_j m_k} - \underset{\substack{m,n\\m\neq n}}{M}\left(\frac{r_{m_j n_j} + r_{m_k n_k}}{2}\right)$	$\underset{\substack{j,k\\j\neq k}}{M}\;\underset{m}{M}\; r_{m_j m_k} - \underset{j}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\; r_{m_j n_j}$
EXTERNAL VALIDITY	Convergent Validity		$\underset{m}{M}\; r_{m\, m_j}$		$\underset{j}{M}\;\underset{m}{M}\; r_{m\, m_j}$
	Discriminant Validity	EDV1	$\underset{m}{M}\; r_{m\, m_j} - \underset{\substack{m,n\\m\neq n}}{M}\left(\frac{r_{m\, n_j} + r_{n\, m_j}}{2}\right)$		$\underset{j}{M}\;\underset{m}{M}\; r_{m\, m_j} - \underset{j}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{r_{m\, n_j} + r_{n\, m_j}}{2}\right)$
		EDV2	$\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{\lvert r_{m\, n_j} - r_{m\, n}\rvert + \lvert r_{n\, m_j} - r_{m\, n}\rvert}{2}\right)$		$\underset{j}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{\lvert r_{m\, n_j} - r_{m\, n}\rvert + \lvert r_{n\, m_j} - r_{m\, n}\rvert}{2}\right)$
		EDV3	$\underset{k}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{\lvert r_{m_j n_k} - r_{m\, n}\rvert + \lvert r_{m_k n_j} - r_{m\, n}\rvert}{2}\right)$	$\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{\lvert r_{m_j n_k} - r_{m\, n}\rvert + \lvert r_{m_k n_j} - r_{m\, n}\rvert}{2}\right)$	$\underset{j}{M}\;\underset{\substack{m,n\\m\neq n}}{M}\left(\frac{\lvert r_{m_j n_k} - r_{m\, n}\rvert + \lvert r_{m_k n_j} - r_{m\, n}\rvert}{2}\right)$

Table 10

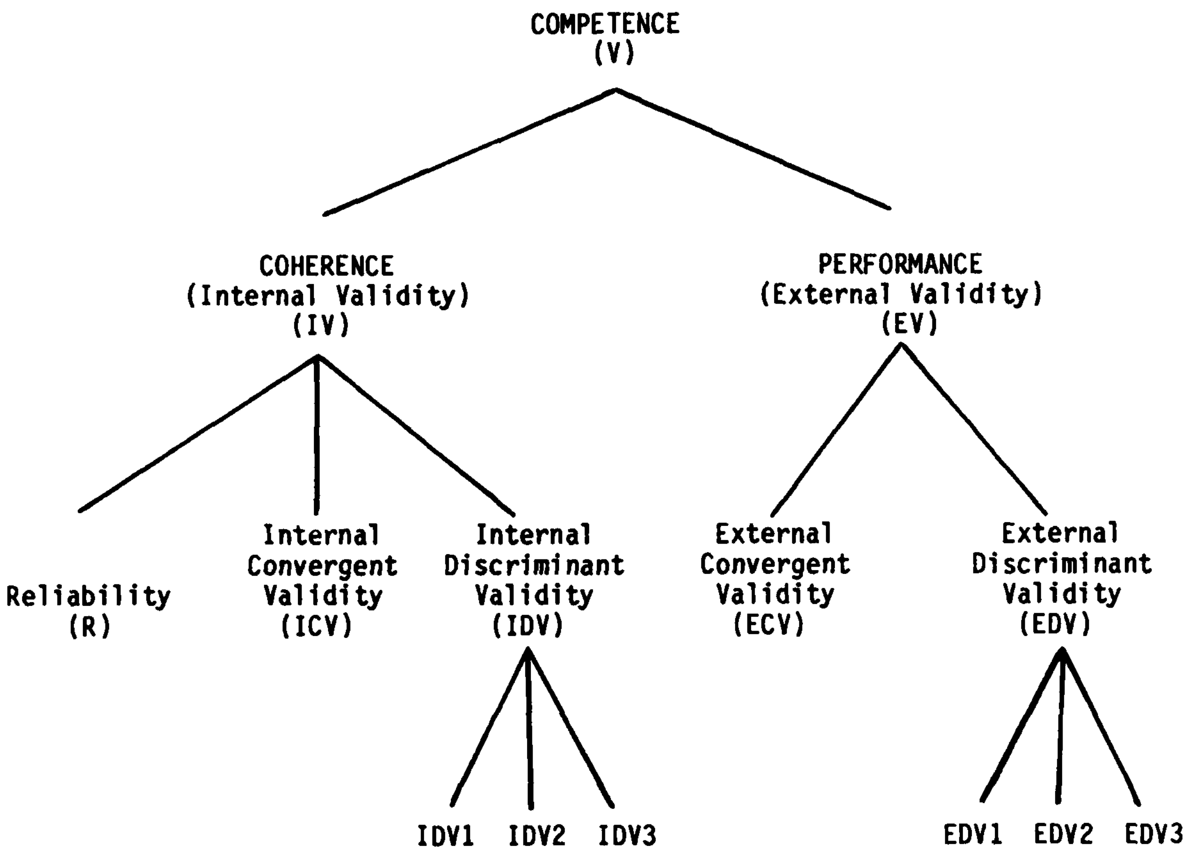
Comparison of Validity Indices (at Overall Level of Aggregation) for Artificial Engineer and Distribution of Validity Indices for Individual Engineers

		Distribution of Individual Engineers		
Index	Artificial Engineer	Low	Mean	High
Overall Validity (V)	.676	.371	.498	.706
Internal Validity (IV)	.853	.403	.619	.886
Reliability (R)	1.134	.706	.961	1.209
Convergent Validity (ICV)	.924	.346	.584	.877
Discriminant Validity (IDV)	.781	.432	.655	.894
External Validity (EV)	.499	.214	.376	.527
Convergent Validity (ECV)	.955	.510	.751	.953
Discriminant Validity (EDV)	.042	-.083	.001	.101

Figure Captions

Figure 1. The structure of indices representing coherence, performance and overall competence.

Figure A-1. Design of the highway engineers study.


COMPETENCE
(V)
COHERENCE
(Internal Validity)
(IV)
PERFORMANCE
(External Validity)
(EV)
Reliability
(R)
Internal
Convergent
Validity
(ICV)
Internal
Discriminant
Validity
(IDV)
External
Convergent
Validity
(ECV)
External
Discriminant
Validity
(EDV)
IDV1
IDV2
IDV3
EDV1
EDV2
EDV3

		METHOD		
		Film Strip (Intuition Inducing) I	Bar Graph (Quasi Rationality Inducing) Q	Formula (Analysis Inducing) A
CONCEPT	Aesthetics E			
	Safety S			
	Capacity C			

APPENDIX A

Context of Application

Whereas Campbell and Fiske (1959) directed their efforts toward ascertaining the validity of measures of constructs ("traits") about people, we attempted in a study of experts to ascertain the validity of expert judgments of concepts about highways. The purpose of this study was to examine the relative efficacy of intuitive, quasi-rational and analytical cognition. Twenty engineers judged the aesthetic value, safety, and capacity of 40 highways under three modes of cognition. Each engineer's judgments were studied in each cell of the diagram presented in Figure A-1. Intuition was induced by requiring each expert to judge each concept (aesthetics, safety, capacity) from film strips of 1-3 mile segments of each of the 40 highways. Quasi rationality was induced by requiring each expert to judge each concept from bar graphs that presented the values of nine attributes for each highway. Analytical cognition was induced by requiring each engineer to construct a mathematical formula for each concept. An empirical criterion was available for each concept. The criterion for the aesthetic value of each highway was derived from the mean judgment of 91 citizens who judged the same highway segments by rating the film strips, or by rating or ranking single frames from the film strips. The criterion for safety was the accident rate for each highway segment averaged over 7 years. The criterion for capacity was the figure calculated by using the procedure from the Highway Capacity Manual 1965 (Highway Research Board, 1965). Each expert devoted roughly 20 hours to the nine sessions, each of which was separated by two-week intervals. (See Hammond, Hamm, Grassia & Pearson, 1984, for details.)

Insert Figure A-1 about here

APPENDIX B

Correction for Attenuation

To use the intra-ecological correlations to estimate discriminant validity accurately in the external validity analysis, two new procedures are described:

1. Comparison of each heteroconcept correlation with the corresponding intra-ecological correlation.

2. Comparison of the order of pairwise heteroconcept correlations with the order of intra-ecological correlations.

These procedures risk being in error if the measures involved in one correlation are more noisy than the measures involved in another, because the true correlation of the noisily measured concepts would be underestimated. We would normally correct for such attenuation, using the formula:

$$rc(a,b) = \frac{r(a,b)}{sqrt(r(a,a)*r(b,b))}$$

where r(a,b) is the correlation between the measures of concepts a and b, rc(a,b) is the correlation corrected for attenuation, and r(a,a) is the reliability of the measure of a.

We have not corrected for attenuation in the illustrative analysis we present here because the reliabilities were not measured in the study by Hammond, Hamm, Grassia, and Pearson (1984). Although estimation procedures for the reliabilities of the engineers' judgments were used in creating the internal validity matrix (see Footnote 1), we hesitate to use these estimates in the above formula because the product would be an "estimate of an estimate". Also, the reliability of the criterion measures can not be similarly estimated because, for example, the capacity criterion was produced from a formula and thus has no measurement error, though it might still be "in error" in that the formula could be wrong.

Because of these problems with the measurement of reliability, the comparisons involved in producing external discriminant validity measures 2 through 5 use correlations that have not been corrected for attenuation. What are the possible effects of this?

1. If the amount of noise is identical for the judgments and the criterion measures, there is no problem; if (as is more likely) there is less noise in the criterion measures than in the engineer's judgments, then in testing Hypothesis 8 we will have <u>underestimated</u> the extent to which the engineers underdiscriminate among the concepts. Further, the measures of EDV2 and EDV3 will be especially noisy.

2. If the concepts are judged or measured with equal amounts of noise, then we have no problem in comparing them; if on the other hand one concept is judged or measured with more noise than another, then the comparison of the patterns in the heteroconcept triangles in Hypotheses C1 and C2 may be distorted.

To avoid such problems, it is important in planning research using the multiconcept multimethod methodology to directly measure the reliability of each judgment and each criterion measure, if possible.

APPENDIX C

Further Measures of Internal and External Discriminant Validity

This appendix explains and demonstrates the second, third and fourth measures of internal discriminant validity and the fifth measure of external discriminant validity.

Internal Discriminant Validity

The second measure compares the correlations on the reliability (monoconcept, monomethod) diagonal (see Table 3) with the correlations in the heteroconcept monomethod triangle. The third measure compares the correlations on the validity (monoconcept heteromethod) diagonal with the correlations in the heteroconcept monomethod triangle. The results of these measures with respect to Hypothesis 4 were identical to those determined by the first internal discriminant validity measure: the safety and capacity concept pair was least discriminable.

The fourth internal discriminant validity method, originally suggested by Campbell and Fiske (1959) in the passage quoted above, examines whether the correlations between judgments of different pairs of concepts have the same pattern regardless of the methods used in making the judgments.

Each of the 9 heteroconcept triangles contains correlations between judgments of each of the three possible pairs of concepts: aesthetics and safety (ES), aesthetics and capacity (EC), and safety and capacity (SC). There are six possible ways in which these correlations may be ordered. Similarity of the pattern of correlations in all nine heteroconcept triangles is evidence that for this set of concepts, this set of methods provides discriminant validity. The distribution of the heteroconcept

triangles among these orders can be tested with Chi-square against the expectation that 1.5 triangles would exhibit each of the 6 orders (cf. Delucchi, 1983).

For example, in the internal validity matrix for the artificial engineer (Table 3), there are 4 triangles with correlations in the order SC > ES > EC, 2 triangles with SC > EC > ES, 2 with ES > SC > EC, and 1 with EC > SC > ES. The Chi-square for the artificial engineer is not significant (Chi-squared = 7.667, df = 5, NS), and there is therefore no evidence for discriminant validity with this procedure, for the artificial engineer. For all engineers:

HC1: There is no predominant pattern among the hetero-concept correlations.

The analysis was carried out for each of the 20 engineers. Six engineers deviated significantly from the expected distribution; that is, showed evidence for discriminant validity. Four of these had the order SC > ES > EC.

External Discriminant Validity

The availability of the criterion measures and their intercorrelations allows us to look more directly at the question that was asked in Hypothesis C1 concerning the relative sizes of the correlations in the heteroconcept triangles. We will present this analysis using only data from the internal validity matrix; a parallel analysis could be done with data from the external validity matrix.

The correlation between the aesthetics and capacity criterion measures (.279) is larger than the correlation between aesthetics and safety (.275), which in turn is larger than the correlation between safety and capacity (.180). Accurate discriminant validity would require that this EC > ES > SC pattern occur in each heteroconcept triangle. (Note, however, that since the EC correlation is almost identical to the ES correlation in this particular data set, the ES > EC > SC pattern would also be expected to occur often.) Our hypothesis is:

HC2: Engineers' heteroconcept correlations have the same pattern as the criterion intercorrelations.

The null hypothesis is the same as for Hypothesis C1. To illustrate the analysis of this hypothesis, none of the artificial engineers' heteroconcept triangles exhibited the expected patterns EC > ES > SC or ES > EC > SC. The Chi-square test was used to determine, for each engineer individually, whether significantly more of his nine heteroconcept triangles had the expected pattern EC > ES > SC or its easily confused competitor ES > EC > SC. This is, of course, a more stringent test than for HC1. It was found that for only one engineer was the EC > ES > SC pattern predominant, and even this was not statistically significant. In fact, the reverse patterns were most common -- eight engineers had SC > ES > EC, and 7 had SC > EC > ES.

The fourth measure of internal discriminant validity, applied in testing Hypothesis C1, and the fifth measure of external discriminant validity, applied to Hypothesis C2, did not reveal any evidence for discriminant validity in this study. This contrasts with the findings

using the other discriminant validity measures. Athough an explanation is available (the engineers judge safety and capacity to be more similar to each other than either is to aesthetics, when in fact aesthetics is more closely related to each than they are to each other), still it is clear that putting requirements on the pattern of heteroconcept correlations represents a stricter test of discriminant validity than the other procedures that Campbell and Fiske (1959) suggested for measuring it.

APPENDIX D

Procedure for Producing Indices of Validity

The various indices (e.g., of internal discriminant validity, external validity, or overall validity) are produced by taking the mean of the appropriate subindices (e.g., the first measure of internal discriminant validity, or external convergent validity) according to the pattern illustrated in Figure A-1. Each subindex is produced for each engineer by taking the mean of z-transformed correlations, from specific locations in the internal or external validity matrices, or the mean of the differences between such z-transformed correlations, corresponding to the comparisons that were illustrated above with Hypothesis 1-8. Table 9 displays the formulas for each of the 9 subindices, at each of 6 possible levels of aggregation. For example, the formula for the internal convergent validity index, at the concept level of aggregation, is:

M
j,k   rm m
j ne k

This index is calculated for each concept m. It is the mean, over all pairs of methods j and k where j is different from k, of the z-transformations of rm m , which is the correlation between two judgments of concept m, using method j and method k. The correlations for the <u>external</u> validity matrix are (with one exception) of form rmn ; that is, the correlation between the criterion measure of concept m and the engineer's judgment of concept n using method j. M is used as a "mean" symbol, representing a sum of correlations divided by the number of correlations summed over. The correlations involved in producing all the subindices in this table have been z-transformed.

Once the subindices are calculated as in Table 9, they combined as indicated in Figure A-1. Thus, the mean of the three internal discriminant validity subindices (IDV1, IDV2, and IDV3) is the index for internal discriminant validity (IDV); the mean of IDV and the internal convergent validity index (ICV) is the index for coherence or internal validity (IV); and the mean of IV and the index for performance or external validity (EV) is the index for overall competence (V).

In order that these indices be on a common scale, in which the meanings of the numbers are preserved when they are involved in the arithmetic operations of calculating means and differences, the indices consist only of those measures of reliability, convergent validity, and discriminant validity that are correlations or differences between correlations (after Fisher's z-transformation of the correlations). Therefore the procedures used for testing Hypotheses C1 and C2 (in Appendix C), which are not expressable as correlations, are not included in this index. Further, the second and third external discriminant validity measures used here are the <u>absolute values</u> of the differences between the engineer's heteroconcept correlation and the corresponding criterion intercorrelations (which addresses accuracy), while <u>relative</u> differences were used to test Hypothesis 7 (which addressed the question of over- or underdiscrimination). Finally, note that at some levels of aggregation specific subindices can not be created. For example, it is not possible to measure convergent validity at the level of concept pairs, because convergent validity deals, by definition, with only one concept. Similarly, it is not meaningful to create an index for the external validity of a pair of judgment methods, for the external validity matrix deals with only one judgment at a time. (A measure was possible for EDV3,

however, because it is derived from the internal validity matrix.) This means that the index should not be used for making comparisons between different levels of aggregation.

These indices are useful for a number of purposes. They can, for example, provide measures for evaluating:

1. Individual engineers' ability to <u>discriminate among concepts</u> (use individual IDV or EDV indices at the Overall level of aggregation in Table 9). In the present study, the engineers' individual internal discriminant validity indices range from .432 to .894, and their external discriminant validity indices range from -.083 to .101.

2. How well individual <u>concepts</u> can be judged (use mean V, IV, or EV indices at the Concept level of aggregation in Table 9). In the present study, aesthetics is judged best (internal validity = .93, external validity = .66), safety next (internal validity = .49, external validity = .23), and capacity third (internal validity = .45, external validity = .24).

3. How well <u>pairs of concepts</u> can be discriminated (use IDV or EDV indices at Concept Pair level of aggregation). In the present study, the aesthetics and capacity concepts are just as easily discriminable (IDV = .83, EDV = .07) as the aesthetics and safety concepts (IDV = .82, EDV = .08); safety and capacity are most readily confused (IDV = .32, EDV = -.16).

4. How well specific methods work (use indices at the Method level of aggregation). Both internal and external validity show that analysis is the best method for judging these concepts (internal validity = .80, external validity = .46), quasi rationality next (internal validity = .61, external validity = .45), and intuition third (internal validity = .44, external validity = .23).

5. How well pairs of methods work (use indices at the Method Pair level of aggregation). Consistent with the previous result, in case one wished to use only two of the three methods on a future project, one would choose the quasi-rational and analytical methods (IV = .63, EV = -.21) rather than the intuitive and quasi-rational (IV = .35, EV = -.23) or the intuitive and analytical (IV = .36, EV = -.24) methods.

OFFICE OF NAVAL RESEARCH

Engineering Psychology Group

TECHNICAL REPORTS DISTRIBUTION LIST

OSD

CAPT Paul R. Chatelier
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D. C. 20301

Dr. Dennis Leedom
Office of the Deputy Under Secretary
of Defense (C^3I)
Pentagon
Washington, D. C. 20301

Department of the Navy

Engineering Psychology Group
Office of Naval Research
Code 442 EP
Arlington, VA 22217 (2 cys.)

Aviation & Aerospace Technology
Programs
Code 210
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Communication & Computer Technology
Programs
Code 240
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Physiology & Neuro Biology Programs
Code 441NB
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Department of the Navy

Tactical Development & Evaluation
Support Programs
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Manpower, Personnel & Training
Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mathematics Group
Code 411-MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Group
Code 411-S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 2217

CDR K. Hull
Code 230B
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Department of the Navy

Special Assistant for Marine Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Dr. J. Lester
ONR Detachment
495 Summer Street
Boston, MA 02210

Mr. R. Lawson
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

CDR James Offutt, Officer-in-Charge
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D. C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D. C. 20375

Dr. J. S. Lawson
Naval Electronic Systems Command
NELEX-06T
Washington, D. C. 20360

Dr. Robert E. Conley
Office of Chief of Naval Operations
Command and Control
OP-094H
Washington, D. C. 20350

CDR Thomas Berghage
Naval Health Research Center
San Diego, CA 92152

Department of the Navy

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, D. C. 20350

Dr. Andrew Rechnitzer
Office of the Chief of Naval Operations, OP 952F
Naval Oceanography Division
Washington, D. C. 20350

Combat Control Systems Department
Code 35
Naval Underwater Systems Center
Newport, RI 02840

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Orlando, FL 32813

CDR Norman E. Lane
Code N-7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Ocean Engineering Department
Naval Ocean Systems Center
San Diego, CA 92152

Department of the Navy

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Ross Pepper
Naval Ocean Systems Center
Hawaii Laboratory
P. O. Box 997
Kailua, HI 96734

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D. C. 20375

HQS, U. S. Marine Corps
ATTN: CCA40 (Major Pennell)
Washington, D. C. 20380

Commanding Officer
MCTSSA
Marine Corps Base
Camp Pendleton, CA 92055

Chief, C^3 Division
Development Center
MCDEC
Quantico, VA 22134

Human Factors Technology Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 334A
Washington, D. C. 20361

Department of the Navy

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D. C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 03416
Washington, D. C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 81323
Washington, D. C. 20360

Larry Olmstead
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

Mr. Milon Essoglou
Naval Facilities Engineering Command
R&D Plans and Programs
Code 03T
Hoffman Building II
Alexandria, VA 22332

Captain Robert Biersner
Naval Medical R&D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Department of the Navy

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Commanding Officer
Naval Health Research Center
San Diego, CA 92152

Commander, Naval Air Force,
U. S. Pacific Fleet
ATTN: Dr. James McGrath
Naval Air Station, North Island
San Diego, CA 92135

Navy Personnel Research and
Development Center
Planning & Appraisal Division
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

CDR J. Funaro
Human Factors Engineeing Division
Naval Air Development Center
Warminster, PA 18974

Mr. Stephen Merriman
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

Department of the Navy

Dean of the Academic Departments
U. S. Naval Academy
Annapolis, MD 21402

Dr. S. Schiflett
Human Factors Section
Systems Engineering Test
Directorate
U. S. Naval Air Test Center
Patuxent River, MD 20670

Human Factor Engineering Branch
Naval Ship Research and Development
Center, Annapolis Division
Annapolis, MD 21402

Dr. Harry Crisp
Code N 51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Mr. John Quirk
Naval Coastal Systems Laboratory
Code 712
Panama City, FL 32401

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

Office of the Chief of Naval
Operations (OP-115)
Washington, D. C. 20350

Professor Douglas E. Hunter
Defense Intelligence College
Washington, D. C. 20374

Department of the Army

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D. C. 20310

Department of the Navy

Dr. Edgar M. Johnson
Technical Director
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and
Systems Research Laboratory
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U. S. Army Human Engineering Labs
Aberdeen Proving Ground, MD 21005

Department of the Air Force

U. S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D. C. 20332

AFHRL/LRS TDC
Attn: Susan Ewing
Wright-Patterson AFB, OH 45433

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks Air Force Base, TX 78235

Foreign Addressees

Dr. Daniel Kahneman
University of British Columbia
Department of Psychology
Vancouver, BC V6T 1W5
Canada

Foreign Addressees

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology
Establishment
Teddington, Middlesex TW11 OLN
England

Director, Human Factors Wing
Defence & Civil Institute of
Environmental Medicine
Post Office Box 2000
Downsview, Ontario M3M 3B9
Canada

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF England

Other Government Agencies

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 copies)

Dr. Craig Fields
Director, System Sciences Office
Defense Advanced Research Projects
Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. M. Montemerlo
Human Factors & Simulation
Technology, RTE-6
NASA HQS
Washington, D. C. 20546

Dr. J. Miller
Florida Institute of Oceanography
University of South Florida
St. Petersburg, FL 33701

Other Organizations

Dr. Robert R. Mackie
Human Factors Research Division
Canyon Research Group
5775 Dawson Avenue
Goleta, CA 93017

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. H. McI. Parsons
Human Resources Research Office
300 N. Washington Street
Alexandria, VA 22314

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

Professor Howard Raiffa
Graduate School of Business Administration
Harvard University
Boston, MA 02163

Dr. T. B. Sheridan
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Arthur I. Siegel
Applied Psychological Services, Inc.
404 East Lancaster Street
Wayne, PA 19087

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Harry Snyder
Department of Industrial Engineering
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

Other Organizations

Dr. Ralph Dusek
Administrative Officer
Scientific Affairs Office
American Psychological Association
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Robert T. Hennessy
NAS - National Research Council (COHF)
2101 Constitution Avenue, N. W.
Washington, D. C. 20418

Dr. Amos Freedy
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert C. Williges
Department of Industrial Engineering and OR
Virginia Polytechnic Institute and State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Deborah Boehm-Davis
General Electric Company
Information Systems Programs
1755 Jefferson Davis Highway
Arlington, VA 22202

Dr. Ward Edwards
Director, Social Science Research Institute
University of Southern California
Los Angeles, CA 90007

Dr. Robert Fox
Department of Psychology
Vanderbilt University
Nashville, TN 37240

Other Organizations

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D. C. 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Mr. Edward M. Connelly
Performance Measurement
Associates, Inc.
410 Pine Street, S. E.
Suite 300
Vienna, VA 22180

Professor Michael Athans
Room 35-406
Massachusetts Institute of
Technology
Cambridge, MA 02139

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Other Organizations

Dr. Babur M. Pulat
Department of Industrial Engineering
North Carolina A&T State University
Greensboro, NC 27411

Dr. Lola Lopes
Information Sciences Division
Department of Psychology
University of Wisconsin
Madison, WI 53706

Dr. A. K. Bejczy
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91125

Dr. Stanley N. Roscoe
New Mexico State University
Box 5095
Las Cruces, NM 88003

Joseph G. Wohl
ALPHATECH, Inc.
2 Burlington Executive Center
111 Middlesex Turnpike
Burlington, MA 01803

Dr. Marvin Cohen
Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. William R. Uttal
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109

Dr. William B. Rouse
School of Industrial and Systems
Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Other Organizations

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th Street
Chicago, IL 60637

Dr. Douglas Towne
University of Southern California
Behavioral Technology Laboratory
3716 S. Hope Street
Los Angeles, CA 90007

Dr. David J. Getty
Bolt Beranek & Newman, Inc.
50 Moulton street
Cambridge, MA 02238

Dr. John Payne
Graduate School of Business
Administration
Duke University
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
School of Engineering and
Applied Science
University of Virginia
Charlottesville, VA 22901

Denise Benel
Essex Corporation
333 N. Fairfax Street
Alexandria, VA 22314

Psychological Documents (3 copies)
ATTN: Dr. J. G. Darley
N 565 Elliott Hall
University of Minnesota
Minneapolis, MN 55455

END

FILMED